

SUPPORTING JUSTIFICATION FOR OMB CLEARANCE

OF DATA COLLECTION INSTRUMENTS FOR THE

STUDY OF ENHANCED INSTRUCTION IN AFTER-SCHOOL PROGRAMS

Submitted May 16, 2005

Submitted by

MDRC
16 E. 34th Street
New York, NY 10016

Fred Doolittle
Project Manager



ENHANCED ACADEMIC INSTRUCTION IN AFTER-SCHOOL PROGRAMS

OMB SUBMISSION

Table of Contents

❖ Supporting Statement

- **Exhibit 1: Theory of Action**
- **Exhibit 2: Data Collection Plan**
- **Exhibit 3: Estimated Hour Burden**
- **Exhibit 4: Minimum Detectable Effect Size**

❖ Appendix A: Impact Study Data Collection Instruments

- **Table A: Impact Instrument Content Chart, Explanatory and Outcome Variables**
- **Parent Application Form**
- **Student Follow-up Survey: Reading and Math Versions**
- **Regular School Day Teacher Survey: Reading and Math Versions**
- **Student Records Request**
- **Achievement Tests: Baseline and Follow-up**

❖ Appendix B: Context/Implementation Study Data Collection Instruments

- **Table B: Implementation Instrument Content Chart**
- **Program Attendance Records**
- **After-School Project Program Director Interview**
- **After-School Project “Enhanced” Curriculum Staff Interview**
- **After-School Project “Regular” Academic Program Staff Interview**
- **After-School Project District Coordinator Interview**
- **After-School Academic Staff Survey**
- **After-School Academic Activity Scales Form**

❖ Appendix C: Confidentiality, Consent, and Contact Forms

- **MDRC Confidentiality Pledge**
- **SRM Staff Confidentiality Agreement**
- **Parent Consent Form**
- **Family Contact Info Form**

SUPPORTING STATEMENT

DATA COLLECTION IN THE STUDY OF ENHANCED ACADEMIC INSTRUCTION FOR AFTER-SCHOOL PROGRAMS

A. JUSTIFICATION

This submission requests approval for a data collection plan for a study of enhanced academic instruction in after-school programs. The project is sponsored by the Institute of Education Sciences within the U.S. Department of Education. The study will examine the implementation and impact of new after-school curricula for second through fifth grade reading and math, with study sites typically receiving 21st Century Community Learning Center (21st CCLC) funding. The 21st CCLC program is authorized under Title IV, Part B of the Elementary and Secondary Education Act of 1965 (ESEA), as amended by the No Child Left Behind Act of 2001 (NCLB) (Pub. L. 107-110). Evaluation is one of the allowable activities under the national activities monies and this study is part of the Department's program improvement efforts, as indicated necessary by prior research. The authority for the evaluation and solicitation and collection of records in the system of records is: (1) sections 171(b) and 173 of the Education Sciences Reform Act of 2002 (ESRA) (Pub. L. 107-279)(20 U.S.C. 9561(b) and 9563); and (2) section 4202(a)(2) of the Elementary and Secondary Education Act of 1965 (ESEA), as amended by the No Child Left Behind Act of 2001 (NCLB) (Pub. L. 107-110) (20 U.S.C. 7172(a)(2)).

Data collection for the evaluation will occur in school year 2005-2006 and involve a baseline questionnaire administered to parents, follow-up questionnaires for teachers and students, baseline and follow-up achievement testing of students, collection of attendance data from program records, collection of student records, and field research on the implementation of the new curricula. The evaluation will use a random assignment design to estimate the impact of the new curricula (as compared to the usual services in after-school programs) on students' academic outcomes. This submission requests clearance for all data collection instruments.

1. Circumstances Necessitating the Data Collection

As the pressure for students to meet challenging academic standards grows, parents, principals, and policymakers are increasingly turning their attention to the out-of-school hours as a critical opportunity to help prepare students for the knowledge demands of our evolving world. Indeed, the federal government has been making a substantial investment towards this goal through its 21st Century Community Learning Center (21st CCLC) funding. Though the program has evolved in emphasis, a primary statutory purpose of this federal money is to "provide opportunities for academic enrichment" to help students meet state and local standards in core content areas. In the initial years of

the program, most local after-school programs primarily provided homework assistance¹ and/or offered home-grown academic enrichment activities that were of mixed quality in addition to a broad range of other enrichment and recreational activities.² As the program has matured, programs have been gradually augmenting their academically-oriented offerings. Though there are a growing number of curricula developed for the after-school setting, to our knowledge none have scientifically sound evidence indicating they improve academic outcomes.

Coincident with this programmatic evolution, the Department of Education began a program of research seeking to strengthen the ability of after-school programs to support academic growth and healthy youth development for students. While many aspects of this program of work include efforts to strengthen program operations and to identify promising approaches, the Institute of Education Sciences has also funded evaluations of the effects of after-school programs on academic and other outcomes for children. In the initial study, the research examined the impact of providing elementary school students access to after-school programs (providing the existing mixture of academic support and other services). The experimental research design involved a program group provided access to the after-school program (most of whom got help with homework) compared to a control group not admitted to the program. (In addition, there was a complementary non-experimental study of the impacts of after-school programs for middle school students.) The studies examined impacts on a range of outcomes, but the findings related to academic effects are most relevant here. Both studies found no or small effects on young people's academic performance, as measured by grades and test scores.³ Dynarski and his co-authors found that, "At the elementary school level, reading test scores and grades in most subjects were not higher for program participants....In addition, on average, programs had no impact on whether students completed their homework or completed assignments to their teacher's satisfaction."

Building on this initial set of findings, the Department reasoned that an important next step was to develop stronger instructional resources for core academic subjects (i.e., reading and math) that could be used in after-school programs and test their effectiveness in improving academic outcomes. This study is a result of that effort. It addresses the question whether enhanced academic instruction in after-school programs based on curricula adapted from the regular school day produces better academic outcomes than the typical after-school academic support, often consisting of help with homework or locally assembled materials. The Department views this study as what is sometimes

¹ Dynarski, M., C. Pistorino, M. Moore, T. Silva, J. Mullens, J. Deke, P. Gleason, W. Mansfield, S. James-Burdumy, S. Heavyside, L. Rosenberg, and D. Levy. (2002) *When Schools Stay Open Late: The National Evaluation of the 21st-Century Community Learning Centers Program*. Washington, D.C.: U.S. Department of Education, Office of the Under Secretary.

² Grossman, J.B., M. Price, V. Fellerath, L. Jucovy, L. Kotloff, R. Raley, and K. Walker. (2002) *Multiple Choices After School: Findings From the Extended-Services Schools Initiative*. Philadelphia: Public/Private Ventures and MDRC. See also, Walker, K. E., and A. J. A. Arbreton, with the Stanford University School of Education Research Team. (Forthcoming) *Final Report for the San Francisco Beacons Initiative*. Philadelphia: Public/Private Ventures.

³ Dynarski et al., 2002.

called an *efficacy study* in the evaluation literature: a test of the potential of a new approach to improve outcomes. (Flay, 1986) It is not a standard effectiveness evaluation, seeking to understand how an existing program as typically implemented affects outcomes.

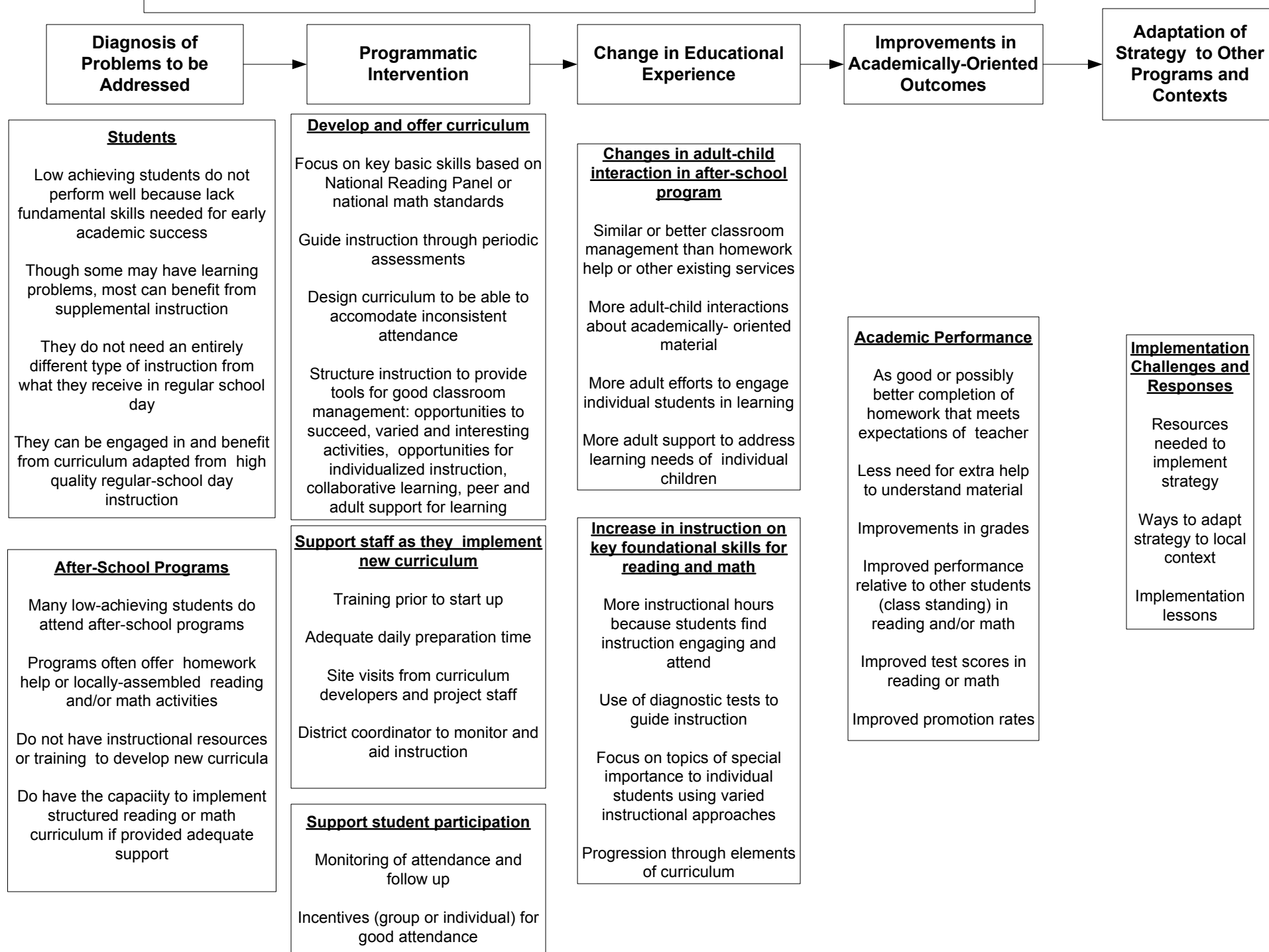
2. How, By Whom, and for What Purpose Are Data to be Collected and Used

The design of this project calls for adaptation of existing reading and math curriculum from the “regular” school day to fit the needs and unique circumstances of the after-school setting. This approach is important for several reasons. First, compared to much of the existing after-school material, “in-school” instructional material tended to be more closely aligned with research evidence the Department felt provided guidance on how best to teach reading and math in the elementary grades. In addition, many in-school curricula are aligned with state standards outlining what children are expected to learn. Further, there was more research suggesting the effectiveness of various in-school curricula in improving academic outcomes than was the case for the after-school material, though much of this research had not used the strongest research designs. Moreover, with the limited time available in after-school programs, it is arguable that approaches offering instruction focused on key skills and thought processes are more likely to affect academic outcomes than instructional approaches relying on projects or more self-directed study. Finally, many organizations providing in-school curriculum have much experience and resources to draw on in adapting instructional material to the after-school setting and could mount the type of effort needed for a major impact study.

Exhibit 1 shows a theory of action underlying the new interventions and the evaluation design. It begins with a diagnosis of the nature of the problems facing students in low performing schools and the capabilities of students and after-school programs to address these problems. It then summarizes the logic underlying the programmatic interventions (new curricula) being developed, implemented, and tested in this project. It then outlines the intended changes in the educational experiences of students receiving the enhanced after-school instruction (through the new curricula) as compared to other students who receive the regular after-school service offerings. The exhibit then summarizes the academic effects hypothesized to occur because of the new interventions. One outcome in this stage of the model (completion of homework) is specified in a different way than others (as good or better completion of homework) because of the way the new curricula are likely to be implemented. The new curricula will be offered as the first activity after snacks (when the control group will typically be getting homework help); in some sites, this schedule may result in the program group getting less help than the control group in completing homework. The hypothesized improvement in academic skills because of the enhanced instruction might be counter-balanced by the decrease in time devoted to homework help. Thus, the hypothesized effect recognizes the possibility homework completion will not actually improve. The theory of action shown in Exhibit 1 concludes by identifying next steps in the possible adaptation of the strategy to new settings if the results are favorable.

a. Key Questions for the Research

Exhibit 1: Theory of Action Underlying the Project and Evaluation



The impact study will focus on several questions:

- Does enhanced instruction in after-school programs using a reading or math curriculum adapted from the regular school day significantly improve reading or math proficiency as measured by test scores?
- What are the effects of the after-school reading and math interventions on other in-school outcomes such as grades and standing relative to other students, homework quality and completion, and retention/promotion?
- Do the interventions' impacts differ for key subgroups of students? Within the constraints of the sample, we will estimate impacts for subgroups based on prior academic performance, race, gender, grade level, and other key theoretically-supported subgroups.

Our goal in the context study is to understand if and how implementation of the enhanced instruction changed the after-school program experience of students. This will support our understanding of the appropriateness of the theory of action, the fidelity of implementation of the enhanced instruction, the service contrast between the program and control groups, and our impact findings. It will also allow us to describe the activities involved in instituting the enhanced instruction for other programs who might wish to implement a similar strategy. These goals lead to a series of questions for the context study:

1. **Context:** What is the local context in the study sites?
 - a. What is the local educational setting and challenges?
 - b. What is the basic organizational structure and staffing of the programs?
 - c. Who are the students served?
 - d. How is the enhanced instruction using the new curriculum fit within the goals, structure, and schedule of the program?
2. **Fidelity:** How strong is the fidelity of implementation of the enhanced instruction using the new curriculum?
 - a. How is the program staffed to provide the new curriculum?
 - b. How well do staff implement the curriculum?
 - c. What proved to be hard?
 - d. How did staff respond?
 - e. What suggestions do they have to improve implementation
3. **Service Contrast:** What is the service contrast between the program group (receiving instruction using the new curricula) and control group (receiving the usual after-school services)?
 - a. How does staffing vary in terms of ratios of students to staff, skills of staff, etc.?
 - b. What is the difference in activities offered, in terms of topics, format, frequency, etc.?

- c. How do the activities vary in terms of general practices in serving children? Are there differences in activity organization and structure, classroom management, appropriate facilities, etc.?
4. **Implementation Lessons:** What are the implementation lessons coming out of the sites that could strengthen the effectiveness of the interventions or support its adaptation to additional programs?

b. Overview of the Project Plan

A guiding principle in developing this project was the objective of structuring a fair test of the concept of enhanced academic instruction based on in-school curriculum. By a fair test, we mean selection of sites able to operate the curricula to be studied, identification of a sample of students who are the intended “target group” for the interventions, creation of a clear and policy relevant service contrast between the program and control groups, a strong research design to provide credible estimates of the impacts of the program (both in methodology and sample size), collection of data on the relevant range of outcomes, and use of implementation and contextual data to interpret the impact findings. The resulting plan for this project involves three key steps:

1. Implementation of the curricula in approximately 25 after-school centers for reading and 25 after-school centers for math during school year 2005-6. These centers must also have an existing or usual mixture of academically-oriented activities that are a clear service contrast with the enhanced curricula and must serve students in need of additional academic support to meet local academic standards;
2. Use of a random assignment impact design to allocate students in the after-school program to receive the enhanced curricula (the program group) or continue to receive the usual after-school academically-oriented services (the control group). This will allow us to assess the impacts of the new enhanced curricula on a range of academically-oriented outcomes with a sample of approximately 2000 students for the test of the reading curriculum (divided evenly between the program and control groups) and 2000 for the test of the math curriculum (again assigned evenly to the program and control group);
3. Analysis of the local context, fidelity of implementation of the curricula, and service contrast between the program and control groups to aid in the interpretation of the impact findings.

This project plan leads to the data collection effort summarized in Exhibit 2, which lists the various purposes for which data will be collected, the data sources, the respondents, how data will be collected, and when data will be collected. The Appendices to this Supporting Statement discuss in detail the various data sources planned for the project.

The planned data collection related to the impact study questions includes:

Exhibit 2
Data Collection Plan for the Project

Purpose	Data Source	Respondent	Person Collecting Data	When Collected
To describe the study sample	Parent /guardian application form	Parent or guardian	After-school program staff includes application as part of enrollment and research team collects form when submitted	As part of after-school program enrollment process, prior to random assignment
	Baseline reading or math achievement test	Student	Research team supervises administration of reading or math achievement test by program staff	At start of after-school program in fall 2005, prior to random assignment
To describe participation in the activities of after-school centers in the study	Program attendance or participation data	After-school program staff	Research team collects program data on attendance in after-school activities	Throughout school year 2005-2006
To describe service contrast between sample members in the program and control group. Key topics include participation in other academically-oriented out-of-school activities and special help students receive in school, which could lessen service contrast if controls get more	Follow-up student survey	Student	Research team fields and collects surveys	Fall 2005 and Spring 2006
	Follow-up survey of sample members' school-day teacher	Regular school-day teacher	Research team fields teacher survey and collects completed surveys.	Spring 2006

Purpose	Data Source	Respondent	Person Collecting Data	When Collected
To describe local context, services offered in the after-school centers (both enhanced instruction and normal services). Allows for assessment of match between curriculum and local needs, fidelity of implementation, service contrast, and implementation lessons.	Field research, interviews, staff survey, observation of activities	Program staff	Research team interviews program staff, administers staff survey, and conducts observations.	Early 2006 field research visit to each center
	Follow-up student survey	Student	Research team fields and collects surveys	Fall 2005 and Spring 2006
To estimate differences in student academic outcomes between the program and control groups	Follow-up reading or math achievement test	Student	Research team administers reading or math achievement test depending on curricula tested in center	Spring 2006
	Follow-up survey of sample members' school day teacher	Regular school-day teacher	Research team fields survey with teachers and collects completed surveys	Spring 2006
	Grades and other measures from student records	School staff	Research team collects school records	Summer/Fall 2006

- **Baseline parent questionnaire** collected by the after school program staff prior to random assignment to describe the study sample and the other resources students can draw and have drawn on for academic help,
- **Baseline achievement testing** conducted by the research team prior to random assignment to describe the existing math or reading achievement level of students in the survey and to form subgroups for the impact analysis based on initial achievement levels,
- **Field research on curriculum implementation** conducted by the research team in early 2006 to understand the context in which the curricula are implemented, how well they are implemented, the service contrast between the program and control group, and implementation lessons,
- **Follow-up student surveys** conducted by the research team in the fall of 2005 and the spring of 2006 to understand students' reactions to the after-school program and other academic supports they may receive,
- **Attendance data for the after-school program** collected by the research team during school year 2005-6 to document participation in the after-school program and the special curricula,
- **Follow-up teacher survey** conducted by the research team in the spring of 2006 to understand students' academic performance relative to other students and any special academic support they receive during the school day,
- **Follow-up achievement testing** conducted by the research team in the spring of 2006 to measure impacts on achievement test scores,
- **Student school records**, collected by the research team in the fall of 2006 to measure impacts on additional academic outcomes.

We plan to rely on the following data collection efforts and analysis strategies to address the four context study questions:

- **Context:** Data from interviews with after-school program staff, local district staff, and document review will be used to characterize the local context in each site. Appendix B includes interview protocols for the local program director, enhanced (new curriculum) instructors, regular academic activity instructors, and the district coordinator supporting implementation of the new curricula. It also includes a survey for after-school staff involved in any academically-oriented activity, both the enhanced instruction using the new curricula and any existing academic support. These data will be used to describe the local context in which the new curricula are being implemented.
- **Fidelity of Implementation:** The study will rely on operational data sources for this analysis. As part of the operation of the curriculum in each site, the local district coordinator supporting implementation in the program will be conducting observations of each classroom. This will be done using a standard form developed for operational purposes that identifies important practices for instruction using each of the two new curricula. The information will be collected to identify teachers providing instruction with strong fidelity (who could model good practices) and others needing additional professional development and

support. The evaluation team will collect write-ups of up to three observations of each classroom in each site. In addition, during a research team site visit in early 2006, a research team member (who has been trained to use the operational observation form) and the local district coordinator will observe a classroom together and independently fill out the operational form. They will then compare ratings and discuss any differences to help the research team understand how the operational staff members define strong implementation, clarify key constructs, and provide the team an independent assessment of the fidelity of implementation of the curriculum. This information will be supplemented by attendance data from existing program records to assess how well sites were able to engage students in instruction and generate strong participation. Our plans for collecting attendance data are summarized in Appendix B. Further, the operational staff will develop summary ratings of the strength of implementation of the curriculum for each classroom that will be shared with the research team. Through discussions and comparisons of these data sources, the research team will be able to develop a rating of the fidelity of implementation at the classroom level in each site.

- ***Service Contrast:*** To address this question, the research team will conduct interviews with local program staff and a survey of staff involved in academically-oriented activities in the after-school program. These are included in Appendix B. We will also draw on the program attendance data discussed in the previous bullet. In addition, the research team will observe academically-oriented activities in the study sites. The After-School Academic Activity Scales Form included in Appendix B provide a way to record general characteristics of these activities. Our plan is to conduct two observations of the enhanced instruction using the new curriculum in each site during our site visit, with the classrooms selected randomly. This will help us characterize general instructional techniques, classroom management, facilities, etc.. From our other research on each site, we will know which sites offer any academically-oriented activity in addition to help with homework. If a site does offer some additional academic support activity, we will conduct one observation of this activity, using the same observational scales in order to describe this activity using the same metric as used for the enhanced instruction. This will facilitate summarizing the service contrast. The After-School Academic Activity Scales Form involve a four point scale, reflecting past experience that providing observers these choices generates more reliable information than forcing a yes/no choice. We anticipate collapsing the categories in the analysis into relatively strong versus relatively weak. The combination of these various data instruments will allow us to allocate sites and classrooms to cells in a two-by-two matrix representing stronger and weaker implementation of the curriculum and stronger and weaker classroom management and general instructional practices.
- ***Implementation Lessons:*** These will be developed drawing on data drawn from all the sources references in Appendix B.

Thus, the data to be collected will be used for multiple purposes, all of which will be realized in the project report completed in 2007. This report will:

- describe the study sample;
- document participation in the special after-school curricula being tested, the usual after-school center activities, and other academically-oriented out-of-school activities;
- describe the implementation of the reading and math curricula being tested and the typical academically-oriented services offered in the after-school programs in the study (which are received by the control group);
- compare the academically-oriented out-of-school service receipt of the program and control groups; and
- estimate differences in student academic outcomes produced by the special curricula.

The project report will address a central issue in after-school policy and practice: whether the strategies of instruction tested here produce improvements in academic outcomes for students in low-performing schools. This is a central issue within the after-school program and policy community, so we anticipate great interest in the report. Many after-school programs are seeking evidence on effective strategies for aiding students academically and there is little solid guidance to follow. With the performance expectations on after-school programs increasingly including academic progress and schools seeking new ways to improve learning and achievement, we anticipate a substantial audience for the findings among program funders and operators. The precise dissemination plan for the report will be developed later in the project, but we anticipate it will include postings on websites, announcement of the study findings, briefings for key staff in the federal government, and presentations at professional meetings.

3. Use of Automated, Electronic, Mechanical, or Other Technological Techniques

The data collection plan seeks to reduce respondent burden by using existing data available in an automated or electronic form when possible. For example, we will rely on attendance data collected by the after-school programs in the study as part of their normal operations to document participation in the special curricula being tested and the normal services offered by the centers. We will give the participating centers the option of submitting these data in a variety of automated formats or in hard copy formats such as attendance rosters. We will follow a similar strategy with our collection of student records listing grades and academic progress (promotion/retention and other variables), giving participating sites several options as to how to collect these data.

However, several key types of data needed for the analysis can only be obtained from parents, students, school-day teachers, and after-school program staff. The baseline and follow-up student achievement tests administered as part of this study (needed to have consistent measures of achievement across the many districts in the research) will be primarily fielded in a group setting, though some short tests of key aspects of reading achievement (fluency, word recognition) will be administered individually. We will utilize the appropriate test form for students in grades two through five and will score the

tests using available scanning technology and software. The follow-up student survey will be administered by on-site members of the research team in the fall of 2005 and the spring of 2006, with as many surveys as possible fielded in a group setting, and telephone and in-person backup for students not completing the survey as part of the group.

With the sample covering grades 2-5, it will typically be possible to identify a single school-day teacher for each sample member who can provide crucial outcome data through a teacher survey. Members of the research team will work with local school staff to prepare the list of the appropriate teachers for this survey and the students for whom each should complete a survey. A member of the research team will provide these teachers with surveys for each of the sample members they teach. A member of the research team will collect completed surveys by mail or other means and follow up with teachers who have not completed surveys.

4. Efforts to Avoid Duplication of Data Collection

The data collection effort planned for this project will produce data that is unique and targets specifically the research questions the Department identified for this project. The services tested have been developed as part of this project, so there are no other programs providing these services where we can rely on existing data for the evaluation. The information collected from parents, students, teachers, and program staff is not available elsewhere and is focused on topics needed to address the project's research questions.

Because we are using an experimental (random assignment) impact design, we are able to focus baseline data collection on items important to describe the sample, define subgroups of students, or increase the precision of impact estimates (covariates in the analysis). Thus, the baseline form completed by parents is relatively short.

In general, we have avoided collecting the same data from multiple sources (which is sometimes done to check accuracy). There are a few exceptions, mostly in the implementation protocols, where capturing multiple perspectives was necessary and justifiable and the number of people interviewed in a site is small.

Students are being asked to complete a survey on their reactions to the after-school program and its academic component at two points in time and to provide information on other academic supports they are receiving outside the regular school day. The survey is very short and we have chosen a fall and spring fielding of the survey to allow these students to focus on their early and later reactions to the after-school program and to lessen any recall problems that could occur if we asked them to think about an extended period of participation.

We are planning to administer an achievement test to students in the study sample at two points in time (baseline and the end of the school year) to allow us to describe the initial academic skills of the sample and estimate impacts on student achievement. We recognize that most schools and districts do administer achievement tests to students as

part of the usual school day program. Nevertheless, we believe a special achievement test is needed for this study for four reasons:

1. **Better precision for the impact estimates.** The existing achievement tests in use in the study sites will vary greatly in their content, design, and reported scores making it difficult or impossible to describe the academic achievement of the sample in any consistent way across sites. Some sites will have state competency-based tests measuring achievement of state-specific standards, while others will be nationally normed tests of academic achievement. Further, the reported scores will vary greatly, with some sites providing raw scores, normal curve equivalents (NCEs), percentiles, grade equivalents, and/or categorical designations with many different definitions (pass, basic, proficient, advanced, etc.). By having a common test at baseline (for use as covariates) and follow-up (as a key outcome), the precision of impact estimates will be improved substantially.
2. **Better measure of learning over a consistent timeframe.** The timing of the existing achievement tests will vary across sites in ways that make their use in the evaluation problematic. For baseline testing, we would most likely typically have to use tests from the prior academic year, though some schools might have fall testing. This could introduce errors into data since students could experience varying learning loss or gains over the summer. In addition, these test scores would not be available for students who are new in school year 2005-6 to the districts in the study. Further, the timing of spring testing will vary considerably meaning students in sites would have had varying exposure to the curriculum at the time they are tested. For example, sites that administer an existing achievement test in late winter or early spring would have much less chance to produce impacts on achievement than sites where the existing test was fielded in April or May.
3. **Lessened chance of missing data.** Students in the research sample may not have existing test scores because they may not be tested because of local or state exemptions from testing, because they were absent when testing occurred, because they entered the study school after school year 2004-2005 testing, and/or because they left the study school prior to the testing in school year 2005-2006. If we were to rely exclusively on student records containing scores from existing tests, we would not know the extent of these missing score issues until we accessed the data in the summer or fall of 2006. At this point, it would be too late to respond to the problem through fielding a test for students without scores.
4. **More timely receipt of data.** Obtaining data from existing testing can be a complicated and lengthy process. In some prior education studies, we have not been able to obtain the final test score data until 6 months or more after the end of a school year. This could seriously delay the completion of the project.

Recognizing that existing achievement tests are likely to be fielded in participating sites, we are planning to administer achievement tests to sample members that are as short as feasible to collect the data needed for the study. The precise choice of tests is not yet final, though discussions with the Department of Education staff and other experts will soon be concluded and a final selection made. The baseline test will be a group-administered test of about one hour or less, covering only the topic addressed in the curriculum to be tested in the site. Our plan is to select a test that is nationally recognized, normed to a national sample of test takers, similar to tests that are part of state and/or local accountability systems (so it will have policy relevance), and relatively easy to administer. At follow up, we will field a similar group-administered test and expect to supplement this in the reading sites with short individually-administered tests to fill in key topics not covered adequately on the group-administered test.

5. Burden on Small Entities

The primary entities that are a part of the study are schools and school-based after-school programs. Burden on these entities has been minimized by focusing the data collection effort on key topics, and streamlining the collection process. Our baseline data collection is limited (as permitted in an experimental design), our teacher and student surveys are quite short (approximately 5 minutes), and the achievement testing zeroes in on content directly related to the curriculum being tested (only reading in the reading sites and only math in the math sites).

Further, our staffing plan for the project includes members of the research team who will work closely with each center and school to minimize the burden on them. In each site, we will have a project-funded on-site data collection coordinator present to work with program staff to lessen the burden of data collection. At key points in the data collection effort (especially follow-up achievement testing), other members of the research team can be on-site to facilitate data collection and lessen the burden on program and school staff.

6. Consequences for Federal Programs or Policies if the Collection is Not Conducted or is Conducted Less Frequently

This study is a one-time data collection effort. It is part of the overall Department of Education effort to improve the quality of the 21st CCLC program by developing and testing new educational resources. The Department is seeking models of instruction proven effective through rigor research that can guide state and local practice.

Without the data and findings from this study, the Department will be unable to test the effectiveness of substantive academic programs produce academic improvements in students in low-performing schools. Given the importance placed on academic improvements in these schools under the federal *No Child Left Behind* legislation and the role of federally-funded after-school programs in helping to produce the needed improvements, the study's findings are central for the development of federal policy.

7. Special Data Collection Circumstances

There are no special circumstances related to the data collection for this project.

8. Federal Register Announcement and Consultation Prior to OMB Submission

a. Federal Register Announcement

Published in *Federal Register*, Volume 70, No.44, Tuesday, March 8, 2005 at page 11197-8.

b. Consultation on Data Collection

The research team has consulted with a project Technical Working Group, made up of the following individuals:

- Megan Beckett, Full Social Scientist, the Rand Corporation
- Tom Dee, Assistant Professor, Department of Economics, Swarthmore College
- Carolyn Denton, Assistant Professor, University of Texas at Austin, Department of Special Education
- Larry V. Hedges, Professor of Education, Psychology and Sociology, University of Chicago
- Nancy C. Jordan, Associate Professor of Education, University of Delaware
- Rebecca Maynard, Professor of Education and Social Policy, University of Pennsylvania

In addition, the research team consulted surveys and data collection forms and protocols used in the following prior studies:

- Evaluation of the 21st Century Community Learning Centers
- Evaluation of Extended Service Schools
- Evaluation of the San Francisco Beacon Initiative
- Parent Observation of Child Adaptation
- Michigan Study of Adolescent and Adult Life Transitions
- NICHD Study of Early Child Care and Youth Development
- Patterns of Adaptive Learning (University of Michigan)
- University of Rochester Self-Determination and Motivation Study

9. Justification for Payments to Respondents

We plan to follow the *Guidelines for Incentives for NCEE Evaluation Studies* prepared by the Department of Education and dated 3/22/2005. This outlines the circumstances in which respondent incentives are appropriate in NCEE studies and the maximum amounts permitted. This study meets the criteria for the use of respondent incentives outlined in these Guidelines. Specifically, the *Guidelines* propose incentives when the following conditions are met:

- a. *The population includes a control group in an important (and perhaps expensive) study where it is imperative to keep most respondents in the control group sample or the result of the whole study could be vitiated.*** This study is a random assignment impact study to which substantial Department of Education resources have been committed. Without high respondent completion rates for both the program and control groups, the investment of Department funds will not produce valid findings.
- b. *The target population (students, teachers, and after-school staff) is a small group that is often surveyed, meaning any particular respondent is liable to be in somebody's sample frequently.*** While not frequently in a sample such as this study, there are many reporting and assessment demands on districts and schools from the federal and state levels, as well as from other researchers seeking to conduct studies. The schools and after-school programs in this study are already under special scrutiny since they were selected because they have many students not yet meeting academic standards. Districts and schools frequently limit researchers access and require time-consuming procedures for approval. Compensating respondents for their time and the added burden can make access easier to obtain. Further, federal and state requirements for testing have increased dramatically in recent years, and will continue to increase in the near future. While this testing is needed for the reasons discussed elsewhere in this statement, the tests planned come on top of existing data collection. Teachers are also the target of numerous requests to complete surveys on a wide variety of topics from state and district offices, independent researchers, and the Department of Education (PPSS and NCES). Further, the teachers' school days are already quite busy, potentially requiring them to complete surveys outside school time. There are also in some localities collective bargaining agreements that do not allow teachers to complete surveys during school time.
- c. *There is a lengthy field period.*** The data collection needed for this study is spread throughout the school year and is fairly intense. For sites in this evaluation, there will be a student pre-test administered in the fall, a student survey in the fall and spring, a teacher survey in the spring, after-school program observations and interviews for multiple days during the year, and a student post-test administered in the spring. With this lengthy and intense fielding period, it is important to maintain a commitment to provide the data needed for the study.
- d. *There are unusual demands or intrusions on the respondent.*** Students will be asked to respond to two brief surveys about their reactions to the after-school programs and other academic support they receive and they will also be asked to take two achievement tests for baseline and follow-up data. This data collection will occur primarily in after-school hours, when such data is not typically collected. The primary strategy is to collect these data within the

after-school program. However, this may not always be possible (some students may stop attending the program). In these instances, it may become necessary to use incentives for those students no longer attending the after-school program. In addition, regular school day teachers will be asked to complete a short survey on students in the sample. This will be unusual in that the students will be part of a study of activities outside the regular school day, rather than during the school day. Some teachers may have only a peripheral or indirect connection to the after-school activity.

- e. ***It is important to convince hard-core refusals to respond, especially in small sub-populations of interest.*** We will be testing and surveying students in our sample who stop attending after-school programs (in order to maintain the integrity of the random assignment design). They must make special arrangements to be tested and to complete surveys. Having incentive payments will be important in gaining high response rates from those no longer participating in the after-school program.

Our data collection plans have been crafted to have the minimal feasible burden, so we believe this study fits within the low burden category on the incentive schedule at the end of the *Guidelines*. We also believe that some of the data collection activities can be undertaken with incentives that are less than the amounts listed in the *Guidelines*. This is especially true for the student assessments. We also believe that incentives will not be necessary for most of the study sample. However, for some of the data collection (e.g., “hard-core” refusals), we think that a minimal level of incentives will be the cost-effective method of attaining adequate response rates required by OMB. Therefore, we will field the data collection plan using the following incentives as ***ceilings or maximum***, with the goal of attaining high response rates with smaller incentives where possible:

Teacher rating of students	\$3 per student
Student survey (for those no longer attending the program)	\$15
Student assessment (for those no longer attending the program)	\$50

For the two last items (student assessments and surveys), we will work to avoid respondent payments but recognize that conducting these for students no longer attending the after-school program may involve the cooperation of parents and may impose some special burdens on them if they cannot be done during the regular school day. Thus, we believe having the option to compensate parents for this burden could be important in producing the needed response rates.

10. Confidentiality of the Data

MDRC and other organizations that are part of the research team will follow procedures for assuring and maintaining confidentiality that are consistent with the provisions of the Privacy Act and other relevant federal requirements. Specifically, data collection activities will be conducted in compliance with The Privacy Act of 1974, P.L. 93-579, 5 USC 552 a; the “Buckley Amendment,” Family Educational and Privacy Act of 1974, 20 USC

1232 g; The Freedom of Information Act, 5 USC 522; and related regulations, including but not limited to: 41 CFR Part 1-1 and 45 CFR Part 5b and, as appropriate, the Federal common rule or ED's final regulations on the protection of human research participants.

Respondents to all data collection instruments will receive information about confidentiality protections at the beginning of a form or outset of the interview. Respondents will be informed that all of the information they provide will be kept strictly confidential and that the results of the study will be presented only in aggregate form.

The following safeguards are routinely employed by MDRC to carry out confidentiality assurances:

- All staff members – at MDRC, P/PV, and Survey Research Management – sign an agreement to abide by the corporate policies on data security and confidentiality. This agreement affirms each individual's understanding of the importance of maintaining data security and confidentiality and abiding by the management and technical procedures that implement these policies. MDRC and Survey Research Management use these forms as a matter of course in their work with research data, and copies are included in Appendix C. The study team members from P/PV will sign an adapted version of the MDRC form.
- All data, both paper files and computerized files, are kept in secure areas. Paper files are stored in locked storage areas with limited access on a need-to-know basis. Computerized files are managed via password control systems to restrict access as well as physically secure the source files.
- Merged data sources have identification data stripped from the individual records or encoded to preclude overt identification of individuals.
- All reports, tables, and printed materials are limited to presentation of aggregate numbers.
- Compilations of individualized data are not provided to participating agencies.
- Confidentiality agreements are executed with any participating research subcontractors and consultants who must obtain access to detailed data files.

Furthermore, the initial package of application materials that parents/guardians will receive so they can sign up their children for the study and the after-school program includes a consent form (see Appendix C) that explains that the study team will keep data about their families confidential. All data collected, quantitative and qualitative, will be used for analytical purposes only.

11. Additional Justification for Sensitive Questions

The data collection instruments largely cover non-controversial subject matter in standard ways. For example, they do not cover topics often considered sensitive such as risk-taking behavior, sexual or political topics, criminal activities, etc.

12. Estimates of Hour Burden of Data Collection

Participation in all data collection activities is completely voluntary, with no sanctions or penalties being applied for respondents who choose not to provide information or who do not answer specific questions. Exhibit 3 summarizes the estimated respondent burden for the various data sources. As summarized in the chart, the estimated respondent burden for data collection is 9600 respondents and 3,201 hours. Not included in this total is the time involved in administering baseline and follow-up achievement tests to the 4,000 students in the research sample (totally approximately 10,000 hours).

13. Estimate of Total Cost Burden to Respondents

There are no costs of the planned data collection for respondents other than the estimated time discussed in the prior section.

14. Estimate of Total Cost to the Federal Government

The estimated cost to the federal government for the entire project – both operational support for sites and evaluation - is approximately \$12,500,000 spread across four years. This constitutes an average cost of approximately \$3,125,000 per year.

15. Reasons for Changes or Adjustments in Burden

As a new data collection effort, this research involves a change of 3,201 hours in data collection burden.

16. Tabulation, Analysis, and Publication Plans and Schedule

The data collected for the study will be used in the following ways.

Describing the Study Sites: The data from the context study will be used to describe the sites in which the study was conducted. This will allow the readers to understand the settings in which the interventions were implemented, which will probably be largely school-based after-school programs. The data used will be largely from document reviews and interviews with local after-school staff and school officials.

Describing the Study Sample: The parent application form, school records information, and baseline achievement test will be used to describe the sample of students in the study. Data collected will include standard demographics, prior academic experience/performance, and academic support children have available in non-school hours. These data will also be used to describe the extent to which the program and

Exhibit 3
Estimated Hour Burden for Data Collection

Instrument or Data Source	Average Number of Respondents	Number of Responses per Respondent	Average Time Per Response	Total Burden (Hours)
Parental application form	4,000	1	10 minutes	667 hours
Follow-up student survey	4,000	2 (Fall 2005 and Spring 2006)	5 minutes	667 hours
After-school staff survey	400 (8 staff per center in 50 centers)	1	10 minutes	67 hours
Regular school-day teacher survey	600 (12 teachers per center [3 per grades in four grades] in 50 centers)	6.7 students per teacher	5 minutes per student	333 hours
After-school program attendance records	50 (one staff person per center)	8 monthly reports	60 minutes per month per center	400 hours
After-school program staff field research	400 (8 staff per center in 50 centers)	1	45 minutes	300 hours
After-school services observation	100 (2 observations per center)	1	60 minutes	100 hours
School records form)	50 (one staff person per center)	1	800 minutes per center with 80 students per center and 10 minutes per student	667 hours
TOTAL	9,600			3,201 HOURS

Note: These totals do not include the time involved in administering baseline and follow-up achievement tests to the 4,000 students in the research sample (approximately 10,000 hours).

control groups created through random assignment are similar (which is very likely with the large samples planned for the study).

Describing the Implementation of the New Curricula: Data from the field research conducted in program sites, after-school staff surveys, and observation of after-school activities will be used to understand the extent to which the new curricula are implemented as intended, issues that arose in implementing the curricula, and responses to these issues. In addition, the operations team on the project will be monitoring and supporting implementation of the curricula and will prepare periodic reports on the strength of implementation and the issues arising and responses. We will draw on the data to develop ratings of the strength of implementation in each site. This analysis will allow us to identify sites where the new curricula are being implemented especially well and to create categories of sites based on implementation fidelity.

Describing Services Received by the Program and Control Groups and the Resulting Service Contrast: The after-school program attendance records, after-school staff surveys, field research, and follow-up student surveys will be used to document the extent to which members of both research groups received academically-oriented services outside of regular school time and the characteristics of those services and will summarize findings by describing the resulting service contrast between the two groups. We anticipate creating categories of sites based on the extent of the service contrast present. This overall analysis of services received and of the service contrast in sites will be important in interpreting the findings on the effects of the new curriculum on academic outcomes.

Estimating the Effects of the New Curricula on Academic Outcomes: The follow-up achievement tests, regular-school day teacher follow-up survey, and student records will be used to calculate academically-oriented outcomes for members of the program and control group. They will then be used to calculate estimates of impacts of the new curricula by comparing outcomes for the program and control group.

Such estimates can be expressed as a program-induced difference of means for continuous outcome measures such as scale scores, or as a program-induced difference in proportions or percentages for binary outcome measures such as whether or not students surmount a particular achievement threshold. In addition, they can be applied to outcomes which focus on student academic achievement, such as standardized test scores, as well as to outcomes which focus on student persistence and advancement, such as promotion rates or retention rates from grade-to-grade. In this analysis, we will include as covariates in the analysis key baseline characteristics of the sample to increase the statistical precision of the estimate.

Further, to the extent that the sample size allows, we will use characteristics of students collected at baseline to define policy relevant and theoretically-based subgroups of the sample and examine impacts on these subgroups. The study has not been designed with the statistical power needed to examine impacts on small subgroups, but we will be able to look at subgroups that account for a substantial portion of the sample.

When estimating the effects of enhanced after-school instruction from an experimental sample the first logical step is to specify an impact model for each center in the sample. The next step is to decide how best to pool findings across centers. Together these two steps define a model for estimating the average effects of enhanced after-school instruction. Such estimates can be expressed as a program-induced difference of means for continuous outcome measures such as scale scores, or as a program-induced difference in proportions or percentages for binary outcome measures such as whether or not students surmount a particular achievement threshold. In addition, they can be applied to outcomes which focus on student academic achievement, such as standardized test scores, as well as to outcomes which focus on student persistence and advancement, such as promotion rates or retention rates from grade-to-grade. In the following discussion we refer to all of these outcome measures as post-tests.

For a Single After-School Center

Equation 1 represents a simple regression-adjusted difference of mean outcomes for the program group students (all applicants randomized to the enhanced instruction) and control students (all applicants randomized to the control group) from a single center. The outcome measure, Z_{2i} , is a student post-test score, the covariate, Z_{1i} , is a student pre-test score and enhanced instruction status is represented by a binary indicator, M_i . To facilitate pooling across centers, all test scores are standardized in terms of their corresponding mean and standard deviation for the center's control group (they are expressed as z-scores).⁴

$$Z_{2i} = \alpha + \beta_0 M_i + \beta_1 Z_{1i} + e_i \quad (1)$$

where:

- Z_{2i} and Z_{1i} = the post-test score and pre-test score for student i , respectively,
- M_i = one if student i were randomized to the enhanced instruction and zero otherwise,
- e_i = a random error term that is independently and identically distributed across students.

The coefficient, B_0 , represents the impact of being randomized to the enhanced instruction instead of to the control group. Thus, it represents the impact of an *offer of access* to the enhanced instruction versus whatever other educational alternatives exist for students.⁵ Because this estimate is based directly on the groups produced by random assignment it benefits fully from the methodological strengths of experiments. However, because some applicants randomized to the enhanced instruction will not attend, B_0

⁴ Under the current research plan, we will be administering the same baseline achievement test in all centers in the study, so the scores will be comparable. This presentation is a more general form of the analysis, in case we were to shift to using existing school district tests.

⁵ Drawing on terminology from medical research, this type of result is often referred to as the impact of “intent to treat.”

understates the impact of *enhanced instruction attendance*. Under certain conditions it is possible to adjust experimental impact estimates for non-attendance (discussed later).

For Multiple After-School Centers

There are two fundamentally different ways to pool estimates of the effects of enhanced instruction across centers. One approach, which is often referred to as a “fixed effect” model, addresses the question: What is the average effect of enhanced instruction for centers in the study sample and is the estimate obtained for this average statistically significant? The other approach, which is often referred to as a “random effect” model, addresses the question: What is the average effect of enhanced instruction for the population of centers that is represented by a study sample, and is the estimate obtained for this average statistically significant? Hence the fixed effect model restricts its inferences and statistical tests to the experimental sample of centers, whereas the random effect model attempts to infer to a broader population of centers. To date, given the typically small number of sites for most social experiments, it has been common practice to use fixed effect models for pooling experimental findings. This is because few sites do not provide enough information about how true impacts vary across sites to support broader generalizations with adequate precision. Further, the goal of this study is to conduct an efficacy study of the effects of a new approach and sites are not selected to be a random sample of a larger population of sites. Instead, they are selected because they can provide a “fair test” of the new curriculum because of their ability to implement the curriculum with reasonable fidelity, the expectation of a clear service contrast between the program and control groups, and their ability to meet the requirements of the research (especially the sample site requirements). Therefore, we will be using a fixed effects model for this analysis.

A multi-center fixed effect counterpart to the single-center impact model in Equation 1 can be expressed as the following ordinary least squares regression model.

$$Z_{2ij} = \sum_k \alpha_k S_{kij} + \beta_0 M_i + \sum_k \beta_k S_{kij} Z_{1ij} + e_{ij} \quad (2)$$

Where:

Z_{2ij} and Z_{1ij} = the post-test score and pre-test score, respectively for student i from center j ,

S_{kij} = an indicator variable for each center in the sample, which equals one if a student is from the center indicated ($j=k$) and zero otherwise,

M_{ij} = one if student i from center j were randomized to enhanced instruction and zero otherwise,

e_{ij} = a random error term that is independently and identically distributed across students.

The coefficient, B_0 , represents the overall average impact of being randomized to enhanced instruction instead of to the control group *for the centers in the experimental sample*. The hypothesis test for this coefficient indicates whether the estimated average impact for the sample of centers is statistically significantly different from zero. This

analysis does not attempt to generalize statistically beyond the observed sample of sites. Nevertheless, when interpreting its findings one should describe the types of sites in the sample and generalize beyond them in a more intuitive way. This heuristic approach is especially appropriate for sites that are chosen in ways that make it difficult to identify or describe the population they represent. By confining statistical inferences to the sites in this case, and making broader generalizations less formally, one can avoid giving the impression of “spurious scientificness”.

Other features of the analysis plan include:

Controlling for pre-tests scores: Our analytic models and calculations of minimum detectable effects presume that baseline or pre-test data will be available for sample members. This information can increase the precision of impact estimates, especially for fixed effect models, because pre-tests substantially reduce random post-test error, which is the only source of uncertainty in a fixed effect model.

Adding other covariates: One can also add other characteristics of students and centers as covariates in the impact estimation model. Here, this is not likely to contribute much to the precision of the analysis beyond the improvement produced by including pre-tests. For student-level covariates, this is the case because they typically add little to the predictive power of pre-tests. For center-level covariates, the main limitation is the number of degrees of freedom available for estimating how they are related to outcomes or impacts for centers.

Estimating impacts for student subgroups: The one application for which pre-random assignment data on student characteristics will be important is the estimation of net impacts for subgroups of students that are defined in terms of these characteristics. In particular, there is a good chance of obtaining adequate precision for net impact estimates for subgroups of students that are represented at all or most of the centers in the study sample. And for some of these subgroups there might be adequate precision for estimating the differences in their net impacts. At this point, we plan to examine impacts for subgroups based on gender and existing levels of reading or math achievement.

Dealing with missing data: The simplest such issues to resolve are for missing pre-test data. This is because unbiased impact estimates can be obtained without any pre-test data given the proposed experimental design. Having pre-test data for some, but not all sample members thus only reduces precision.⁶ We plan to minimize or ideally eliminate missing

⁶ If pre-test data were missing for an entire center, one could estimate its impact separately and then pool this estimate with a composite estimate for all other centers. If pre-test data were missing for some but not all students from a given center, one could estimate the impact separately for students without pre-tests and pool this estimate with that for all other students from the center and all other centers. If pre-test data were missing for some students from many centers, then a straightforward extension of the preceding analysis could be used to combine estimates that use pre-tests with those that do not. This procedure can be implemented in a pooled estimation model by specifying a “missing data dummy variable” and interacting it with the covariate in a way that effectively separates estimates of the pre-test coefficient for students with and without the covariate.

pre-test data through our planned random assignment procedures which involve pretesting of students prior to random assignment.

Missing data for post-tests pose a problem that is more serious and more difficult to solve because it requires omitting sample members from the impact analysis, which can produce selection bias if this attrition is substantial and non-random. Thus, we plan substantial tracking efforts to assure we can locate and test a high percent of students in the spring of 2006, at a point at which attendance in after-school programs may have declined. For the most part, only those who move out of their initial school district before the post-test is administered will be lost to follow-up. In our analysis, we will check to see if there is evidence of response bias in the follow up data collection and – if present – apply weights to the available follow up observations to seek to address it. In our analysis, we do not impute missing follow-up outcome data for members of the sample, but rather exclude them from the analysis. But our first line of defense is to attain a high response rate.

Adjusting for non-attendance in enhanced instruction: Some students who are randomized to enhanced instruction will not attend. Hence the impacts estimated for the full experimental sample represent the average effects of having *access* to enhanced instruction. Although this result has important policy implications in its own right, it understates the average effect of attending enhanced instruction. But if it is reasonable to assume that enhanced instruction has no direct effect on offerees who do not attend at all, then it is possible to adjust the experimental estimates of the impact of access to enhanced instruction to obtain “semi-experimental” estimates of the impact of enhanced instruction attendance. In the simplest possible case this adjustment is obtained by dividing the experimental estimate of the impact of the offer of enhanced instruction by the proportion of offerees who actually attend.⁷ Whether or not the conditions for this adjustment are likely to be met in practice remains to be determined. And whether or not it makes conceptual sense to use the adjustment depends in large part on the nature of the alternative educational options available to members of the control group. In a school district with few or no other options that are comparable to the enhanced instruction being studied, it would be more appropriate to consider making the adjustment in the hopes of providing at least an approximate measure of the impact of enhanced instruction attendance. But in a district with many similar options, where many control members may be served by similar types of educational programs, then making the adjustment might have limited utility. Of course, in such environments there may be little utility to conducting the experiment in the first place. Thus, we will seek to avoid such districts in our site selection process.

⁷ For a discussion of this issue see Howard S. Bloom (1984) “Accounting for No-Shows in Experimental Evaluation Designs, *Evaluation Review*, Vol. 8, No. 2, 225-246. The approach described in Bloom (1984) represents a special case of the application of instrumental variables estimation to experimental data. For a more general discussion of this issue see Joshua Angrist, Guido Imbens and Don Rubin (1996) “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association* (91).

Exploratory Analysis of the Relationship between Post-Random Assignment Characteristics and Impacts: To the extent that groupings of sites emerge in the course of project implementation (for example, greater or lesser fidelity of implementation of the curriculum) we will use field research data to explore how fidelity is associated with impacts. In addition, if there are clear subgroups of students emerging based on the extent of participation or the extent of the service contrast, we will explore the link between these factors and impacts.

Publication Plans and Schedule: The project schedule is as follow:

- Project design from October 2003 to summer 2005,
- Selection of sites by summer 2005,
- Implementation of the new curriculum in the impact study sites and data collection in school year 2005-2006,
- Completion of data collection and creation of analysis files by December 2006,
- Analysis of findings in January through March 2007, and
- Preparation of a project report for review by the Department and release in August 2007.

17. Approval Not to Display the Expiration Date of OMB Approval

Not applicable. We are not seeking this and plan to display the expiration of OMB approval on data collection forms.

18. Exception to the Certification Statement

Not applicable. We are not seeking exceptions to the certification statement.

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

The evaluation of enhanced academic instruction in after-school programs will be conducted in a purposive selection of after-school programs which provide an appropriate setting in which to conduct an efficacy test of the new curricula developed for the project. Because of this site selection goal and the resulting site selection process, we will not generalize findings to a broader sample of after-school programs from which the study sites were selected, will clearly state this in the project reports, and will use a fixed effects model for estimating impacts (as discussed above).

In order to provide a fair test of the new curricula, we will include after-school programs:

- serving the intended target group of students who are low-income, attending low performing schools, and not currently meeting academic standards,
- willing and able to implement either the new reading or math curricula with reasonable fidelity,
- offering existing academic support that is a clear contrast with the curricula to be tested (most likely homework help and locally-developed or assembled academic materials that are not a real curriculum), and
- willing and able to follow the research procedures and contribute at least 60-80 2nd - through-5th graders for the research sample.

Our process for identifying study sites involves several steps:

1. Work with the Department of Education and Learning Point Associates (which maintains the 21st CCLC database) to develop a list of all 21st CCLC grantees,
2. Select grantees operating elementary school programs,
3. Supplement this list with other possible after-school programs identified through other contacts (national organizations, other research networks, states, etc.), and
4. Screen this list to identify programs with the needed size (number of students), target group (students not meeting academic standards), capacity to operate the new curriculum, service contrast, and interest in and ability to being part of the study.

We anticipate that most after-school programs in the study will be 21st CCLC grantees, but we may include a few other appropriate programs not receiving this federal funding.

Within this general guideline and process, we anticipate including approximately 25 after-school programs for the test of the reading curriculum and a similar number for the test of the math curriculum. Our site selection process involves broad notices to 21st CCLC grantees operating programs at elementary schools, state 21st CCLC program coordinators, national organizations of after-school programs, regional consortia of

schools, and urban districts. Our initial material describes the study and nature of the curricula being tested. Subsequent material and discussions with potential sites will outline the operational procedures for implementing the curricula in the study, the responsibilities of the study team and participating sites, and the details of the research design. While the goal of conducting a fair test of the curricula is paramount, we anticipate enrolling sites from various regions of the country, and from large, medium, and small cities.

2. Statistical Methods for Sample Selection and Analysis

Identification of the research sample will occur in two stages. First, we will identify a sample of after-school programs for the study, as described in the previous section. Within each after-school program in the study, students identified by local program staff as in need of supplemental academic assistance will be randomly assigned to either the program group (which is provided access to the enhanced curriculum) or to a control group, which will receive the usual academic support offered by the after-school program. As part of the sample intake and random assignment process, parents or guardians of students will complete an informed consent form, a study application form, and a contact sheet (to facilitate fielding of the parental follow-up survey). In addition, students in the research sample will take a short achievement test in reading (for centers testing the reading curriculum) or math (for centers testing the math curriculum). We will work with study sites to incorporate these data collection steps into the application process for the after-school program prior to random assignment, and anticipate that the information collected in these data collection steps will be available for every member of the study sample.

The proposed research plan involves a sample of approximately 80 students in each participating after-school program, with this sample randomly assigned with equal probabilities to the program and control group. This design will generate the following research sample:

- 1000 Reading curriculum program group members
- 1000 Reading curriculum control group members

- 1000 Math curriculum program group members
- 1000 Math curriculum control group members

Follow-up data on student outcomes will be collected through a follow-up achievement test (in either reading or math), student follow-up surveys, a survey of the regular school-day teachers of sample members, and student academic records from their schools. Our study design anticipates that the response or completion rates on these data sources will be approximately 85 percent or greater.

Exhibit 4 shows the minimum detectable effects for the planned sample and number of sites for each curriculum test. Within the separate studies of reading and math, we will be able to statistically detect a 3.0 percentile difference in test scores between program

EXHIBIT 4

MINIMUM DETECTABLE EFFECT SIZE

Minimum Detectable Difference		
Outcomes	Sample Size	
	<i>1000 Program/ 1000 Control</i>	<i>500 Program/ 500 Control</i>
Effect Size	.10	.15
Example of continuous outcome: Test Score Percentile (0-100)	3.0 percentile difference	4.3 percentile difference
Example of binomial outcome: Student completes homework to teacher's standards often (yes/no)	5.3 % difference	7.4 % difference

Note: These effect size calculations assume:

- a) half the sample of 80 students in each of 25 after-school centers are randomized to the control group status and half to the program group,
- b) a 5% level of significance and a 80% power criteria,
- c) 15% sample attrition at follow-up,
- d) regression analysis using covariates collected at baseline controls for 40% of the outcome variance, and
- e) fixed effects modeling.

group students and control group students with 80 percent power (i.e., 80 percent of the time). Thus, if the average control group student scores in the 60th percentile on our reading or math test, we would be able to detect that the enhanced instructional program improved test scores if the average program student scores in the 63 percentile. Similarly if 55.3 percent of the program group students often submit homework meeting the school-day teachers' standards as compared to 50% of the control group (an 5.3 percentage point difference), we would be able to detect it statistically 80 percent of the time. These impacts represent effect sizes of 0.10. If we wished to investigate the impact of a curriculum on half the sample, such as on girls or on the younger (second and third grade) students, the minimum effect size we could detect reliably would be 0.15 (a 4.3 percentile test score differential or a 7.4 percentage point difference in binomial outcomes).

To put these effect sizes into context, we reference Kane's review of the most recent evaluations on after-school programs where he presented information on how much students in the Los Angeles school district gain in math and reading over an entire school year.⁸ He found that fifth graders gain .32 of a standard deviation in reading and .49 of a standard deviation in math. If participation in the enhanced after-school program can increase a student's reading or math instruction by 60 percent—say from 5 hours a week to 8 hours a week—and these after-school hours are equally productive to in-school hours, we could expect effect sizes of .19 in reading and .29 in math.

3. Methods to Maximize Response Rates and Deal with Nonresponse

We anticipate that 100 percent of the sample will have the parent/guardian baseline application form and baseline achievement test scores collected prior to random assignment. We do expect there will be some *item* nonresponse on the parent/guardian application form despite planned efforts to keep even this low. Our earlier discussion of the analysis plan described how any item nonresponse will be handled in the analysis.

For follow-up data collection, we plan a variety of methods to produce a high response rate. The first crucial step is locating intended respondents. We will be collecting contact information on parents/guardians at intake so we will have leads should families move during the school year. Further, on the baseline application form we will ask parents about the likelihood of a move during the coming school year, in order to identify sample members where extra tracking efforts could be important. In addition, we will seek access to school district records and other databases to locate any families that move. We thus anticipate being able to locate and contact a very high percent of students. As to regular school day teachers, our goal is to survey the primary teacher instructing students in the spring of 2006. We will use school and school district records to identify these intended respondents.

The second step is to induce the intended respondents to provide the desired data. Prior to the desired date for data collection, we will send advance notices to intended

⁸ Thomas Kane, 2004. *The Impact of After-School Programs: Interpreting the Results of Four Recent Evaluations*. A Working Paper for the W.T. Grant Foundation available at www.wtgrantfoundation.org.

respondents telling them of the nature of the data request and the importance of responding. Further, our staffing plan includes members of the research team in each study school district, so we will be able to follow leads on respondents, encourage participation, handle any logistical problems, and boost response rates. For both the student surveys and the follow-up achievement test our plan calls for group administration in a convenient location (probably the after-school center), makeup dates and individual follow-up if necessary. The proposed short supplemental tests of reading fluency and word recognition would be individually administered in the after-school setting. If feasible, we will also organize a group administration opportunity for the follow-up survey of regular school day teachers, but we anticipate this is likely to be done by individual teachers at a location and time of their own choosing.

4. Test of Procedures or Methods

In designing the data collection instruments for this study, we relied heavily on instruments previously used successfully in prior studies. An earlier section of this statement lists studies we consulted in developing the instruments, many of which have been used in prior studies of after-school programs. Thus, we know that many items have already been used in large and diverse samples without difficulties. Because we or other researchers have successfully used virtually all data elements proposed in this submission, we have not conducted detailed pretests on individual items. We have conducted small-scale pretests (involving fewer than nine respondents) to determine the approximate length of time for completion of the instruments and the results of this are reflected in our earlier discussion of respondent burden.

5. Individuals Consulted on Statistical Aspects of the Design

In addition to the members of the Technical Working Group listed earlier in this statement, the following individuals were consulted on the statistical aspects of the planned evaluation of Enhanced Instruction for After-school Programs:

Howard Bloom
Senior Social Scientist
MDRC
212-532-3200

Jean Grossman
Vice President
Public/Private Ventures
Professor
Princeton University
609-258-6974

James Kemple
Director of the K-12 Education Policy Area
MDRC
212-532-3200